Optically Mapping Multiple Bacterial Genomes Simultaneously in a Single Run

Matthew C. Riley[1,2], James Eric Lee[1], Emil Lesho[1,2,3], Benjamin C. Kirkup, Jr.[1,2,3]*

1 Walter Reed Army Institute of Research, Silver Spring, Maryland, United States of America, 2 Multidrug Resistant Organism Repository and Surveillance Network (MRSN), US Army, Silver Spring, Maryland, United States of America, 3 Uniformed Services University of the Health Sciences, Silver Spring, Maryland, United States of America

Abstract

Optical mapping of bacterial chromosomes provides an unambiguous low-resolution sequence scaffold of the entire chromosome. In comparison to some techniques, such as pulse field gel electrophoresis, cost and throughput limit the application of this technique outside of genome finishing. We have demonstrated the production of multiple bacterial maps using a single set of consumables; this significantly reduces the time and expense of map production.

Introduction

Optical mapping is increasingly used to characterize genomes [1,2,3]. The map is an ordered restriction fragment list tracking the physical structure of each replicon without amplification. This is accomplished by binding the DNA physically to a surface (referred to as the 'mapcard'), cutting it at restriction endonuclease sites, and labeling it for visualization. The images are recorded and the length of the ordered fragments is calculated. Because DNA shears during the extraction process, optical mapping is not as simple as following a single chromosome; instead, fragments are assembled together in a process analogous to that used for genome assembly after sequencing. Maps can be used for finding large insertions, deletions, duplications and rearrangements otherwise difficult to characterize even with sequencing; and for scaffolding sequenced genomes without the use of numerous PCR reactions (for example, in conjunction with SOMA; www.cbcb.umd.edu/soma).

Service laboratories charge thousands of dollars for a single map. The instrumentation costs ~$300,000, and the consumables are ~$500 for each run. The time to image the DNA and assemble the map is also significant (~8 per day, per instrument and technician). Simultaneous mapping of multiple genomes would reduce cost and labor while increasing throughput, almost proportionally to the number of maps produced from each run.

There are at least two ways to map multiple genomes on a single mapcard. One is to subdivide the mapcard physically; at least, the portion to which DNA adheres (Figure 1). This subdivides the surface area available for each DNA preparation but otherwise leaves the mapping process unaffected. The preparations are physically separated and do not interfere with each other. This method has been successfully implemented within our group but is limited by the physical ability to subdivide the surface; it seems limited to division in two at this time. It works well because the

total capacity of the surface is often far beyond what is needed to assemble a robust bacterial genome map. The other strategy is to combine the DNA preparations prior to applying them to a single surface. This second method is the discussed at greater length for the remainder of this paper, as it does not have the genome limit of two the other method does. Also, there are implications for potentially mapping mixed unknown samples and resolving at least parts of each genome, if not full length, for identification. It has not escaped our notice that the methods can themselves be combined.

In the context of eukaryotic genomes, mapping multiple chromosomes simultaneously on the same mapcard is inevitably necessary because the DNA from the separate chromosomes cannot be segregated prior to mapping. The mapping of eukaryotic genomes has been accomplished many times. In the case of bacteria, the risk of failure has discouraged attempting more efficient mapping, particularly in closely related organisms. However, as presented here, mapping several bacterial genomes on a single mapcard is practical.

To successfully map multiple bacterial genomes on a single mapcard, the genomes being mapped must be amenable to processing with the same restriction endonuclease. This unavoidable restriction on the selection of genomes to be co-processed is not difficult to overcome. Compatible organisms can be selected in advance. It is this limitation that most clearly enforces the need for some upstream identification of the bacteria to be mapped.

Second, maps might become comingled during assembly of mixed DNA preparations. This would theoretically result from co-mapping genomes which share substantially identical sequence over long stretches. The question is: how long of a stretch, in what organizational pattern? Most stretches of shared homology are too short to cause any trouble. Typical molecule lengths are 150 Kb–1.5 Mb. Coverage depths exceed 40× (frequently 20 k–30 k molecules are imaged). In this context, two 5 Mb bacterial genomes will not be comingled unless they merge into a single

Report Documentation Page		*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **OCT 2011**	2. REPORT TYPE	3. DATES COVERED **00-00-2011 to 00-00-2011**
4. TITLE AND SUBTITLE **Optically Mapping Multiple Bacterial Genomes Simultaneously in a Single Run**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Walter Reed Army Institute of Research,503 Robert Grant Avenue,Silver Spring,MD,20910-7500**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Optical mapping of bacterial chromosomes provides an unambiguous low-resolution sequence scaffold of the entire chromosome. In comparison to some techniques, such as pulse field gel electrophoresis, cost and throughput limit the application of this technique outside of genome finishing. We have demonstrated the production of multiple bacterial maps using a single set of consumables; this significantly reduces the time and expense of map production.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**3**	

Standard Form 298 (Rev. 8-98)
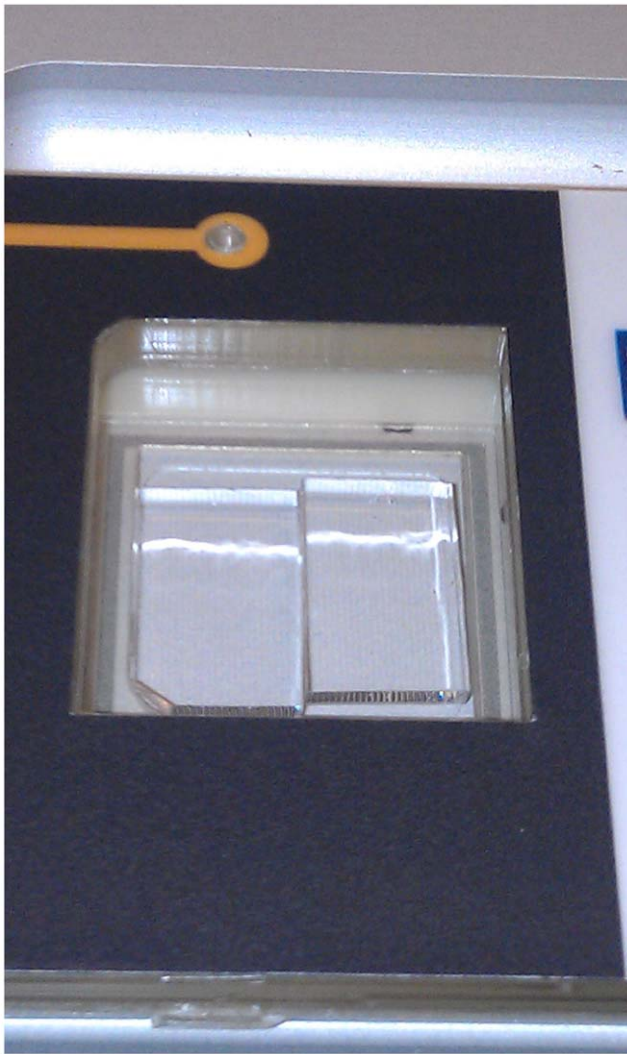Prescribed by ANSI Std Z39-18

Figure 1. Multiple Surfaces. A surface subdivided in two permits parallel mapping efforts during a single run with a single set of consumables.
doi:10.1371/journal.pone.0027085.g001

genome within the limits of certainty for the detection of cut sites over the span of the longest molecules included in the assembly. Even if the assembled molecules were only 300 Kb in median length, continuous stretches of nearly ¼ the genome must be indistinguishable by mapping for concatenation to occur.

As a result, mapping genomes simultaneously will not result in 'false' maps unless two of the strains are nearly identical not just in base pair composition but also in sequence orientation. We have demonstrated mapping of *Shigella dysenteriae* and *Escherichia coli* simultaneously, despite their very close phylogenetic relationship (*Shigella* and *Escherichia coli* are generally considered to be within a single species, but are segregated at the genus level for historical reasons [4]); two clones of *Shigella* would likely not map together successfully using the mixed DNA method. Similarly, based on reference maps being generated at this time, genomes from distinct species of *Acinetobacter* should map simultaneously without ambiguity despite being within a single genus; but many clinical strains of *Acinetobacter baumannii*, distinguished by a few insertions, are too closely related to be mapped independently on the same mapcard.

A third concern is that the quality and condition of one DNA preparation could disrupt reading the entire mapcard; either because the preparation is contaminated or because it is too concentrated. In that second case, the ratio of the DNA preparations results in one map crowding out others. It should be noted that the risk here is merely a waste of material; there is no risk of erroneous data being accepted. Mapping typically involves a quality control step prior to preparing the mapcard. Normal quality control eliminates the risk of contaminating the mapcard or crowding out other genomes by failing poor DNA preparations and ensuring the DNA quality, concentration and length are optimal for a successful preparation; however, in an environment of routine successful DNA preparation, costly quality control steps might be bypassed and the risk of a failed DNA preparation considered acceptable. Ultimately, the combined mapping strategy is more cost and time effective than the single mapping strategy unless a high percentage (~25%, if 3 strains are being mapped together routinely) of the DNA is bad, in which case DNA quality control should be reinstituted.

Results and Discussion

We have already mapped two bacterial genomes on a single mapcard using mixed DNA sample preparations (Figures 2 and 3). Additionally, two *Enterococcus* species were mapped on a single mapcard by cutting the channel forming device into two separate pieces; these have also been mapped independently and are 100% identical as well. Given the data produced from a single mapcard under ideal conditions, this method will permit up to ten bacterial genomes (5 Mb each) to be mapped on a single mapcard. However, the complexity of strain selection and the risks of mapcard failure steadily increase and the benefits diminish as the number of strains per mapcard increases. Three genomes per mapcard represents a substantial cost and time savings; however, the impact of these savings on operations depends on the economics of the laboratory as a whole. Our data (Table 1) indicate that under normal laboratory conditions, three to four strains per mapcard is a reasonable expectation when mixing

Figure 2. *E. coli* Genomes. Complete circularized *E. coli* genome maps created as a result of a mixed DNA mapcard run aligned to an identical map from individual run.
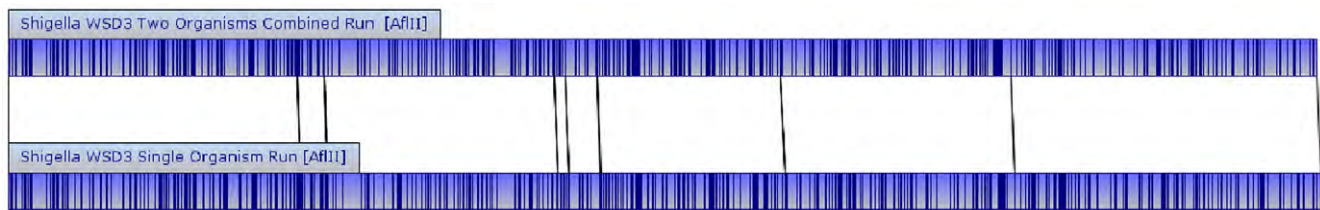doi:10.1371/journal.pone.0027085.g002

Figure 3. *Shigella* **Genomes.** Complete circularized *S. dysenteriae* genomes from mixed DNA mapcard run aligned to genome from individual run.
doi:10.1371/journal.pone.0027085.g003

Table 1. Mapping Statistics.

	Mixed	*Escherichia*	*Shigella*
Total number of molecules:	12,482	7,009	9,804
Genomes Assembled:	2	1	1

These summary statistics list the number of molecules assembled to maps and the number of map assemblies generated during the processing of data from a single mapcard. The mixed mapcard bore DNA from both *Escherichia* and *Shigella*. The maps which were generated were circular and passed standard quality control.
doi:10.1371/journal.pone.0027085.t001

Table 2. Assembly Statistics.

	Aligned maps	Length (Mb)	Avg. depth	Avg. Mol. Size(kb)
E. coli (combined)	1609	5.000	109	341
E. coli (single)	1045	5.160	77	382
S. dysenteriae (combined)	1589	4.176	111	293
S. dysenteriae (single)	1229	4.251	111	386

The columns display the values for the assembly statistics for E. coli and S. dysenteriae genomic maps from the combined run and individual runs. All genomes were circularized and passed standard QC.
doi:10.1371/journal.pone.0027085.t002

DNA preparations. Concerns of lowered quality or depth of coverage are legitimate, but our work here shows (Table 2) it is possible to maintain within the generally accepted scaling factor and far above $40\times$ minimum depth of coverage.

The optimum number of maps on a single mapcard is a function of the operational and economic constraints of each laboratory. At this time, however, we do not foresee a significant gain beyond four maps on a single mapcard, which brings the consumables cost to $129 USD and increases the throughput of the entire process almost 4-fold, given that DNA preparations can be done in parallel. This modification makes mapping process more practical and amenable to incorporation into clinical and near-real time epidemiologic applications.

Materials and Methods

An OpGen® Argus® system maintained under service contract was used to collect the data. The assemblies and figures were created with MapManager™ software. The currently recommended high molecular weight DNA extraction was performed, using magnetic bead technology developed by Agencourt Bioscience Corp®. The protocol provided by the manufacturer was followed unmodified for these samples, though certain organisms require tailored DNA preparation strategies. DNA extraction was performed independently on each strain. Purified DNA from the two strains was diluted to the appropriate concentration (approximately 5 molecules per image at or above ~200 Kb). Each DNA sample was quality checked on a QCard, as was the combined DNA preparation.

Acknowledgments

Author Contributions

Conceived and designed the experiments: BCK MCR JEL. Performed the experiments: MCR JEL. Analyzed the data: MCR. Contributed reagents/materials/analysis tools: EL. Wrote the paper: BCK.

References

1. Latreille P, Norton S, Goldman BS, Henkhause J, Miller N, et al. (2007) Optical mapping as a routine tool in bacterial genome sequencing. BMC Genomics 4(1): 321.
2. Nagarajan N, Timothy D. Read, Mihai Pop (2008) Scaffolding and validation of bacterial genome assemblies using optical restriction maps. Bioinformatics 24(10): 1229–1235.
3. Wellcome Trust Purchases OpGen's Optical Mapping System for Genome Sequence Assembly (2011) Genetic Engineering and Biotechnology News 31(3).
4. Janda JM, Abbott SL (2005) The Enterobacteria, 2nd ed. Washington: ASM Press.